The Statistical Problems of Forensic DNA Analysis

An Honors Thesis (HONORS 499)

by

Michael C. Dart

Thesis Advisor

Dr. Michael P. Brown

Ball State University

Muncie, Indiana

December 1997

The technology of Deoxyribonucleic Acid (DNA) fingerprinting has revolutionized the criminal justice field. It has provided a new tool to convict perpetrators who would previously have been able to go free due to a lack of evidence. There have also been a number of convicted felons who have been released after DNA testing has conclusively shown that they were not guilty. From a single drop of blood, scientists can analyze and compare frequencies of certain DNA patterns, then either include or exclude a particular suspect from consideration.

The major failing of DNA profiling thus far has come in the courtroom when statistical probabilities are cited to a jury to explain the extent to which a suspect is likely to be guilty of a crime. These statistics have been calculated by many different methods, but have not yet displayed that their level of validity is up to the task of deciding guilt or innocence. In this paper, many of the current problems with population genetics and their function in the criminal justice system will be examined and discussed.

Approximately ten percent of human DNA consists of specific coding genes that either contain instructions for the manufacture of proteins or exert a regulatory function. The other ninety percent of human DNA has no known function, and is typically referred to as noncoding DNA (some researchers have also called it "junk DNA"). This noncoding DNA is used in forensic analysis because of its highly variable nature and tendency to mutate between generations.

Coding DNA, on the other hand, has built-in cellular defenses to prevent mutational changes because the slightest change in functional DNA would have a devastating effect on the organism. Researchers estimate that it would take two hundred thousand years to altar a single amino acid in a four hundred amino acid chain mutationally[1], whereas noncoding DNA can change over a single generation.[2]

Another key feature of noncoding DNA is that it contains repetitive sequences of "building blocks". Approximately ten percent of noncoding DNA consists of these tandem repetitive sequences. These repetitive sequences are the source of restriction fragment length polymorphisms (referred to hereafter as RFLPs). RFLPs (pronounced "rif-lips") are the short segments that are produced when DNA is cut with restriction enzymes. Due to the hypervariability of these segments, no two people who are not identical twins have the same sequence of base pairs in their DNA. This allows for a forensic identification with potential equal to that of fingerprinting.

Each RFLP contains a highly variable sequence of nucleotides that have randomly repeating sequences. These variable numbers of tandem repeats (referred to hereafter as VNTRs) can range from a few to several hundred at a single locus. The product of this hypervariability across millions

---

[1] Harding., *et al.*, *Genetics*, v. 132, November 12, 1992, p. 848.

[2] Burke, T., *et al.*, *The Police Chief*, October 1989, p. 92.

Coding DNA, on the other hand, has built-in cellular defenses to prevent mutational changes because the slightest change in functional DNA would have a devastating effect on the organism. Researchers estimate that it would take two hundred thousand years to altar a single amino acid in a four hundred amino acid chain mutationally[1], whereas noncoding DNA can change over a single generation.[2]

Another key feature of noncoding DNA is that it contains repetitive sequences of "building blocks". Approximately ten percent of noncoding DNA consists of these tandem repetitive sequences. These repetitive sequences are the source of restriction fragment length polymorphisms (referred to hereafter as RFLPs). RFLPs (pronounced "rif-lips") are the short segments that are produced when DNA is cut with restriction enzymes. Due to the hypervariability of these segments, no two people who are not identical twins have the same sequence of base pairs in their DNA. This allows for a forensic identification with potential equal to that of fingerprinting.

Each RFLP contains a highly variable sequence of nucleotides that have randomly repeating sequences. These variable numbers of tandem repeats (referred to hereafter as VNTRs) can range from a few to several hundred at a single locus. The product of this hypervariability across millions

---

[1] Harding., *et al.*, *Genetics*, v. 132, November 12, 1992, p. 848.

[2] Burke, T., *et al.*, *The Police Chief*, October 1989, p. 92.

ratios that are usually reported in court to be the "odds" of the suspect committing the crime. The reliability of these "odds" ratios is based on a principle assumption of the Hardy-Weinberg equilibrium - that is, within all allelic frequencies, at each loci, are independent of and have no effect on other loci.[5] The problem arises when Hardy-Weinberg cannot be tested at certain sites in noncoding DNA.

The assumption of statistical independence is based on random mating throughout large homogeneous populations. Scientists claim that the United States is a departure from Hardy-Weinberg since this country has been a genealogical melting pot since its inception, and mating tends to be random within ethnic subpopulations - but not the overall population.[6] The ethnic variations from years of immigration has resulted in a heterogeneous population composed of ethnically and racially diverse groups such as Caucasians, Blacks, and Hispanics. There are further subgroups within each of these major groupings as well. These subgroups are not entirely distinct though, as matings among members of different subgroups do cccur.[7] This becomes an important issue when applied to the courtroom context because different alleles are found to be more frequent in some

---

[5] Cohen, J., *et al.*, *Science*, v. 253, August 30, 1991, p. 1037.

[6] Ibid., p. 1039.

[7] Harding, R., *et al.*, *Genetics*, v. 132, November 1992, p. 855.

subpopulations than others, thus affecting the certainty by which the guilty can be identified.

But is this enough to reject the statistical application of the Hardy-Weinberg principle of statistical independence? According to an article by Devlin (August 1991), statistical deviation from Hardy-Weinberg predictions resulting from subpopulation admixture, can occur only when two criteria are met: (I) there is limited mating among subgroups and (II) there are differences in allele frequencies across the subgroups. Because of the nature of human populations, it should never be assumed that such populations absolutely conform to the principles of Hardy-Weinberg. However, these principles can be used as an excellent approximation to the actual genotype frequencies because of geneflow (intermarriage) among subpopulations and only modest differences, at best, in allele frequencies among subgroups.[8] In fact, these assumptions have been verified repeatedly in human genetics, where a vast array of conventional genetic markers show no deviation from Hardy-Weinberg.[9]

Excess of Homozygosity

---

[8] Cohen, J., *et al., Science,* v. 253, August 30, 1991, p. 1039.

[9] Roberts, L., *Science,* v. 254, December 20, 1991, pp. 1722-1723.

VNTR loci pose difficulty in the assessment of Hardy-Weinberg because of measurement error and coalescence. Improper treatment of such loci can lead to an incorrect conclusion of homozygosity excess. At each VNTR locus, there are a very large number of alleles present. Much of the allelic variation is generated by variation in the number of short, repeated sequences of base pairs linked in tandem in the core region of the locus, which leads to fragment length variation on electrophoresis. The likelihood that each person has unique sequences of VNTRs has made these regions the focus of forensic DNA profiling. However, it has been asserted in the scientific community that several VNTR loci cannot be assumed under Hardy-Weinberg principles because there is an excess of homozygotes at these loci.[10]

A population is in Hardy-Weinberg equilibrium for a particular locus if the probability of observing a genotype (a pair of alleles) in the population is the product of the probabilities of observing each of the pair of alleles in the population when the alleles are the same (for a homozygote)or twice this product when the alleles are distinct (for a heterozygote). Many forces could cause a population to deviate from the assumption of Hardy-Weinberg, such as selection, phenotypic assortative mating, and

---

[10] Lewontin, R., *et al.*, *Science*, v. 254, December 20, 1991, p. 1746.

population subdivision. For VNTR loci in human populations, it has been argued that population subdivision is the most important of these forces.[11]

A population is considered subdivided if there are two or more groups within the population whose individuals experience limited intergroup mating. If the groups differ in their allele frequencies at a given locus and the sampling of the population ignores substructure, an excess of homozygosity will be apparent in the sample even if the subpopulations are in Hardy Weinberg Equilibrium. In this case, the violation of statistical independence is due to the population subdivision, and the probability of observing a homozygote in the population is not the product of the frequencies of its alleles. On the contrary, the degree of homozygote excess depends upon the magnitude of allele frequency differences among such groups, as well as the admixture proportions.[12]

Testing the assumption of Hardy-Weinberg for a population is relatively straightforward. The usual tests involve examining the magnitude of the difference between the observed and expected number of each distinct genotype in a population sampled. Testing Hardy-Weinberg for VNTR loci is more complicated, however, because the length of a

---

[11] Devlin, B., *et al.*, *Science*, v. 249, September 21, 1990, pp. 1416-1418.

[12] Harding, R., *et al.*, *Genetics*, v. 132, November 1992, p. 855.

restriction fragment cannot be measured without substantial error.[13] This error is often larger than the difference in the size of the alleles. Consequently, it is impossible to accurately assign discrete labels to alleles of a VNTR locus. Even if the discrete labels could be assigned, the large number of alleles and genotypes at VNTR loci seriously compromises the power of the usual tests.[14] These difficulties have led to ad hoc methods of analysis of Hardy-Weinberg and contradictory claims. In these ad hoc tests, scientists have three objectives: (I) show that there is an apparent but not real excess of homozygotes at VNTR loci, making previous tests of Hardy-Weinberg invalid; (II) develop an appropriate method of testing Hardy-Weinberg for VNTR loci; and (III)demonstrate the there is no evidence that Hardy-Weinberg is violated for the VNTR loci used in the forensic DNA profiling.[15]

For the tests of excess homozygosity, geneticists have analyzed data from three loci (D2S44, D14S13, D17S79) generated by Lifecodes Corporation for paternity testing and forensic inference. Each locus yields two restriction fragments per individual when cut with the restriction enzyme Pst I. There are three major sources of variation in the measured sizes of restriction fragments. The first

---

[13] Landers, R., *Editorial Research Reports*, v. 1, n24, June 30, 1989, p. 357.

[14] Devlin, B., *et al.*, *Science*, v. 249, September 21, 1990, p. 1417.

[15] Ibid., pp. 1417-1418.

source is intrinsic variation, due to differences in the number of tandem repeats, the size of the flanking region, or both. A second source of variation is measurement error, which can be quite large. (Lifecodes Corporation estimates the error to be $0.006L$, where L is the length of the fragment.) The third source of variation results when an individual's pair of restriction fragments are visualized on x-ray film. Each fragment appears as a distinct band if the pair are substantially different in length (heterozygotes), but the bands are indistinguishable if the pair are the same length (homozygotes). However, for the heterozygotes where the fragments are similar in size but not identical, only one band may be apparent because the distinct bands blur together or coalesce. These heterozygotes are indistinguishable from homozygotes; such a genotype is called a pseudohomozygote.[16]

If coalescence were ignored, or if its consequences were not clearly understood, one might incorrectly infer that there is an excess of true homozygotes in the population. Indeed, the evidence usually cited to reject the assumption of Hardy-Weinberg for VNTR loci is homozygote excess. Coalescence affects the kinds of phenotypes observed in the sample, resulting in an excess of apparent homozygotes and a dearth of "close heterozygotes"

---

[16] Weir, B., *Genetics*, v. 130, April 1992, p. 826.

(heterozygotes with similar allele sizes). Therefore, if a substantial number of observations are affected, this should be readily detectable.[17]

The difficulty of formulating a test of Hardy-Weinberg for VNTR loci had led to some confusion and ad hoc methods of analysis among many microbiologists. These ad hoc procedures treat the apparent homozygotes as if they were all true homozygotes and examine the magnitude of the difference between the observed and expected number of these homozygotes in the sample. These procedures have resulted in contradictory claims about Hardy-Weinberg equilibrium. It is illogical, however to focus on the number of apparent homozygotes in the sample because this number represents an indistinguishable mixture of true homozygotes. If this contaminated data is used without adjusting for coalescence, then any claim that there is no excess of homozygotes in the sample is false. Moreover, for contaminated data, there will be an excess of apparent homozygotes, making the real claim of homozygote excess irrelevant.[18]

In forensic application, where the probability of a multilocus phenotype must be inferred, the usual approach is to assume multiplicability of allele frequencies, both within and across loci. The justification for

---

[17] Devlin, B., *et al., Science,* v. 249, September 21, 1990, p. 1417.

[18] Devlin, B., *et al., Science,* v. 130, April 1992, p. 826.

multiplicability is random mating and Hardy-Weinberg equilibrium. It has been argued that a homozygote excess implies the absence of Hardy-Weinberg, nullifying the validity of multiplication across loci as well as within loci.[19] Although these ad hoc methods do not prove multiplicability across loci, they do suggest that the arguments presented against it are tenuous at best.

Matching Criteria

Another problem frequently cited against the validity of DNA profiling is the match criteria or the reliability of the match/binning process when samples are not easily distinguishable. This system is primarily attacked by statistical geneticists for not taking the statistical effects into full account. One expert in statistical genetics wrote "As far as we know, there were not statisticians involved in developing match/binning. This is hardly the first time that science has taken a wrong turn for lack of guidance from statisticians."[20] In the same vein, another scholar commented that "It is unconscionable that the three laboratories that provide almost all of these analyses have not yet seen fit to employ statistical help in

---

[19] Lewontin, R., *et al., Science,* v. 254, December 20, 1991, p. 1749.

[20] Berry, D., *et al., Applied Statistics,* v. 41, 1992, p. 514.

improving their procedures."[21] Despite these complaints, laboratories still perform their own independence tests and perform to the standards set by the FBI lab. The main obstacle has been getting the validity of these tests accepted in court over the vehement protests of statisticians (typically as expert witnesses).

The use of highly variable marker systems for identification poses a dilemma. The more variants there are at a locus, the less likely it is that two random members of a population will match. For VNTR systems in which each locus has 20 variants, a profile based on four loci has 1.9×10^9 variants while one with eight loci would have 3.8×10^{18} variants, and this is likely to allow almost certain identification.[22] Numerical analysis on forensic databases has confirmed the very large number of DNA profiles in United States populations. One prominent population geneticist has said "The DNA profiling is so highly discriminating that in the event of a match the accuracy with which one estimates the population frequency is of minor importance."[23] Be that as it may, the dilemma arises because of the difficulty of distinguishing between all the

---

[21] Geisser, S., *Journal of American Statisticians Association,* v. 87, 1992, p. 613.

[22] Weir, B., *Proceedings of the National Academy of Science of the U.S.,* v. 89, December 1992, p. 11655.

[23] Evett, I., *Nature,* (London) v. 354, p. 114.

variants at each locus. There are many experimental reasons why it is not possible to infer the number of repeat units with certainty. For example, fragments of lengths 2000 base pairs (bp) and 2020 bp may not be separated on an electrophoretic gel, even if they may represent 200 and 202 copies of a repeat length 10 bp.[24]

The simplest analyses rest on "match/binning," which supposes that the true length of a fragment of estimated length x is enclosed by the interval $x \pm \delta$. Two fragments that have overlapping intervals are said to match; otherwise they do not match. It is generally accepted[25] that standard deviations of measurement errors are approximately proportional to fragment length: the FBI uses a δ value of 2.5% of x, based on empirical studies involving repeated measurements of the same material[26], while Lifecodes Corporation uses a figure of $\delta = 1.8\%$ of x[27], and Cellmark Diagnostics chooses δ values corresponding to a gel migration distance of one millimeter.[28] However, once

---

[24] Weir, B., *Proceedings of the National Academy of Science of the U.S.*, v. 89, December 1992, p. 11656.

[25] Lehrman, S., *Nature, v. 370,* August 25, 1994, p. 588.

[26] Brookfield, J., *Nature,* v. 369, June 2, 1994, p. 352.

[27] Weir, B., *Genetics,* v. 130, April 1992, p. 827.

[28] Ibid., p. 827.

existence of experimental error is recognized, it ceases to be possible to speak in terms of certainty.

Two fragments can be said not to differ in length with a specified level of confidence if their appropriated confidence intervals overlap. An acceptable confidence level is needed, and the distribution of errors must be determined empirically. Measurement errors are correlated, so that determinations of matches for different pairs of fragments are not independent events.[29] Once a match has been declared, the frequency of the matching fragments is found from all of the entries in a database that fall into a "bin" surrounding that fragment length. Care is needed in setting up criteria with appropriate correspondence for the matching and binning procedure. Matching using DNA connected with a particular crime, with the attendant problems of possible contamination and degradation, uses intragel comparisons. Binning, using DNA collected from blood under controlled conditions for establishing databases, is based on intergel comparisons.[30]

Profile Frequencies

---

[29] Weir, B., *Proceedings of the National Academy of Science of the U.S.*, v. 89, December 1992, p. 11655.

[30] Alcamo, I., *DNA Technology: The Awesome Skill*, Dubuque IA, Wm. C. Brown Publishers, 1996, p. 199.

Determining the frequency of a profile found to match between crime scene material and the person of interest is also a statistical issue, since the DNA profiles of all relevant people are not generally available. An estimate of the frequency in the population is needed. The basic "counting" method is not very helpful, since it is very unlikely that any particular profile based on several loci will ever be seen in any sample.[31] Although there is some value in simply presenting the fact that a profile has not been seen in one thousand people examined to date for example, a more informative way of presenting the same information is to say that nonoccurrence in a sample of one thousand implies that it is 99% likely that the true frequency is less than 1 in 218.[32] This is preferable to the confidence limits sometimes used by the defense as estimates of actual profile frequencies, which deflects attention from the fact that the same answer follows from nonappearance of a profile in a sample whether that profile is based on one or several loci. Counting estimates ignore the fact that a DNA profile consists of several components. Figures such as 1 in 218 are seen as misleading when it is realized that, even if everyone in the world had the same two parents, who were heterozygous for different alleles at four independent

---

[31] Lewontin, R., *et al., Science,* v. 254, December 20, 1991, p. 1746.

[32] Weir, B., *Proceedings of the National Academy of Science of the U.S.*, v. 89, December 1992, p. 11654.

loci, frequency of any particular four locus profile would be 1 in 256.[33] A better method of estimating frequencies is needed.

The match/binning procedure estimates profile frequencies as the product of the frequencies of the components of the profile. At one locus, this rests on the Hardy-Weinberg law of population genetics. Much time in court has been wasted by a failure both of the prosecution to establish the independence of allelic frequencies at a particular locus from their databases and of the defense to show that it does not apply.[34] Early trials did not consider any empirical evidence on the issue, and in other cases analogies were drawn to analyses with other genetic markers. Matters were not helped by the repeated reference to tests for consistency of total homozygosity with Hardy-Weinberg expectations.[35] It was not total homozygosity that was at issue. The problem was that population geneticists had little experience in testing for Hardy-Weinberg equilibrium for loci with many alleles, and that the databases contained few of the possible combinations of alleles even at a single loci. It is only recently that demonstrations of Hardy-Weinberg equilibrium in forensic databases have been

---

[33] Weir, B., *Proceedings of the National Academy of Science of the U.S.*, v. 89, December 1992, p. 11655.

[34] Nowak, R., *Science*, v. 265, September 2, 1994, p. 1354.

[35] Devlin, B., *Science*, v. 249, September 21, 1990, p. 1416.

published. Other analyses, based on sampling approaches to exact tests or on shuffling methods, are also showing overall consistency with the Hardy-Weinberg expectations.

The difficulties of establishing independence between pairs of alleles at one locus are magnified when independence between all *2m* alleles for *m* loci are needed. Although the scientific community is coming around to accept the Hardy-Weinberg assumption as well as linkage equilibrium, it will most likely take more published analyses before prosecutors can prove the independence of allele frequencies at loci on different chromosomes.

Population Structure

One situation known to lead to dependence, or disequilibrium, between alleles within and across loci is known as Wahlund effect. It results whenever the population sampled consists of a number of subpopulations with different allelic frequencies. Even if there is independence within each subpopulation, when frequencies are calculated at the population level, disequilibrium will result.[36] Simply put, the product of average is not the same as the average of products.

---

[36] Weir, B., *Proceedings of the National Academy of Science of the U.S.*, v. 89, December 1992, p. 11656.

Opponents of current practices argue that substructuring invalidates use of the product rule, while proponents argue that the issue should not be whether there is substructuring in the population, but whether any substructuring has an appreciable effect on forensic calculations.[37] Neither side has been blessed with much data, and indeed an abundance of this data may prove to be unattainable. The argument for the effects of substructuring that has received the most prominence contends that if two individuals match at one or two loci, this constitutes evidence that they belong to the same population and so will be likely to match at additional loci.[38] The probability of a match is therefore higher than if the individuals (the perpetrator and the suspect) were drawn randomly from the entire population. There may be cases in which there is prior knowledge that perpetrator and suspect do indeed come from the same subpopulation, and then population-wide estimates could be misleading. The counter to this argument has been that the data does not support the notion that such effects will be of forensic significance.[39] True probabilities may be over- or underestimated by two orders

---

[37] Brookfield, J., *et al.*, *Nature*, v. 369, June 2, 1994, p. 351.

[38] Weir, B., *Proceedings of the National Academy of Science of the U.S.*, v. 89, December 1992, p. 11656.

[39] Weir, B., *Genetics*, v. 130, April 1992, p. 886.

of magnitude, but this is unlikely to prevent the probability of a chance match at several loci remaining very small. Estimates of population subdivision parameters have been small within national populations.

Lewontin and Hartl provided an argument of why the U.S. Caucasian population for example, can not be regarded as being genetically homogeneous. This population is derived from genetically diverse European subpopulations in recent generations and, at least initially, tended to maintain this historical separateness in marriage.[40] Taking this argument to its logical conclusion, they say "each particular person may require a different reference group composed of appropriate ethnic or geographic subpopulations." This appears to be heading toward the error made in a recent court decision[41] that said frequency calculations should be based on the distinctive ethnic background of the defendant, who happened to be of mixed Italian, French, and American Indian descent who has pled not guilty, and for whom the ethnic details are unknown. Calculations are necessarily based on random people and are essentially averages over all possible backgrounds."[42] In a paper by Chakraborty and Kidd,

---

[40] Lewontin, R., *et al., Science,* v. 254, December 20, 1991, p. 1748.

[41] Weir, B., *et al., Human Genetics*, V. 50, 1992, p. 869.

[42] Weir, B., *Proceedings of the National Academy of Science of the U.S.*, v. 89, December 1992, p. 11656.

it pointed out that " 'binned allele frequencies' are unbiased estimates, of the averages of all underlying ethnic or endogamous subgroups contained within the reference population"[43] The claim that the ethnic background of the defendant is relevant is one of the most persistent fallacies in the debate, even though its fallaciousness has been well documented.

Calculations of frequencies of a DNA profile in a population are made to quantify the likelihood that some other person left the crime scene material. No calculations are performed if the defendant admits responsibility for the material. Given the defendant's claim, it is necessary to consider the collection of people who could have left the material. There are may be evidence, such as an eyewitness, pointing to a particular ethnic group, or it may be reasonable to simply consider a geographic location. The ideal data would result from a census of the population of possible perpetrators, as when blood samples were taken from 4583 men during an investigation of two rape/murders in Leicestershire and the perpetrator was identified uniquely.[44] The next best data would be a representative random sample of that population.

---

[43] Chakraborty, R., *et al.*, *Science*, v. 254, December 20, 1991, p. 1737.

[44] Weir, B., *Proceedings of the National Academy of Science of the U.S.*, v. 89, December 1992, p. 11655.

Current calculations are based on samples from paternity casework (e.g., Lifecodes Corporation) or from blood banks (e.g., FBI and Cellmark Diagnostics). Another source being investigated is umbilical cord blood samples from maternity hospitals. Paternity data tend to be from a wide geographic area, and blood bank or hospital data from a narrow region. There have been arguments abut the ethnic composition of people becoming involved in paternity disputes or in donating blood, and these are unlikely to be resolved. A counter to the charge of biased samples is to increase the sampling frame by having samples collected from several states and coordinated by the Technical Working Group on DNA Analysis and Methods (TWGDAM). This increased sampling is geographically based, rather than being targeted on specific ethnic groups that might have distinct allele frequencies or, worse, unique alleles.

Continuous Analysis

Many of the population and statistical genetic problems have arisen because of the attempt to apply discrete data techniques to data that are essentially continuous. Estimated fragment lengths do not fall into distinct classes, and the match/binning approach must therefore approximate. A specialized area of research has been

developed making just this point. Unfortunately, it has had no impact on forensic practice in the United States.[45]

In the U.S., continuous analyses are based on the distribution of measurement errors for fragment lengths. Whether or not these errors are normally distributed has been debated, but that is a question that is fairly easy to answer empirically. Joint distributions of errors for two or more fragments can also be found empirically and used in calculations that do not require assumptions of statistical independence.[46]

As with match/binning procedure, however, this still becomes problematic for sets of many alleles. Once the continuous viewpoint is adopted, it is no longer relevant to speak of matching or mismatching. Instead, probability levels can be attached to the joint patterns and a jury could draw its conclusions.

Results from continuous analyses are often expressed as likelihood ratios, in statements such as "The evidence that the crime scene material and the suspect both have the particular DNA profile seen is one million times more likely to have arisen if the suspect provided the crime scene material than if some random unrelated person provided the material."[47]

---

[45] Lehrman, S., *Nature*, v. 370, August 25, 1994, p. 588.

[46] Lewontin, R., *et al.*, *Nature*, v. 372, December 1, 1994, pp. 399-400.

[47] Macilwain, C., *Nature*, v. 381, May 9, 1996, p. 103.

Another use of the likelihood ratio has been proposed by Evett and Werrett[48] as a means of conveying the strength of the evidence to a jury. These authors would say that any likelihood ratio over one thousand constitutes "very strong" evidence, with lesser values having corresponding terms, down to "weak" for ratios between one and thirty-three. This approach appears to avoid arguments over frequencies. Whether or not frequencies may be wrong by a factor of one hundred does not matter beyond the "very strong" limit. The same point was made by Lewontin and Hartl[49] from a different perspective: "After all, 0.0001 is already a pretty small number. Why invoke unsupported assumptions to obtain a still smaller number that is exaggerated and unreliable?" In the end, however, there may be little difference between quoting frequencies or verbal descriptions if the defense asks the prosecution exactly what is meant by "very strong" or if the prosecution claims that there is little significance to a difference in frequencies that are either 10^{-8} or 10^{-10}.

The match/binning approach expresses the final result in terms of P, the population frequency of the matching profile, and $1/P$ is the ratio of probability of the evidence if some unrelated person left the material. If this false positive rate is α, then the likelihood ratio is diminished

---

[48] Evett, I., *Proceedings of the International Symposium on Human Identification*, (Promega, Madison, WI), 1990, p. 89.

[49] Lewontin, R., *et al.*, *Science*, v. 254, December 20, 1991, pp. 1749-1750.

to $(1-\alpha)$ / $(P+\alpha-2P\alpha)$.[50] For $\alpha = 0.001$, the likelihood ratio when P is 1 in 1,000,000 would change from 1,000,000 to about 1000. There has been some debate on the actual error rates in proficiency tests, and there is a strong movement to implement better tests.

The five statistical problems associated with using forensic DNA analysis in the courtroom: difficulty accounting for excess homozygosity (whether actual or perceived); unacceptable matching criteria; unsuccessful analysis of profile frequencies; difficulties assessing population structure; and the wrongful application of discrete data techniques to a continuous analysis, have all been thoroughly debated in the scientific community. Every day new methods for circumventing or disproving these problems are developing. But beyond these scientific problems are perhaps even more salient issues, such as the difficulty in explaining these technologies to judges and jurors. However, until the microbiologists and population geneticists can see eye to eye, DNA in the courtroom will continue to be a stalemate battle of expert witnesses.

---

[50] Weir, B., *Proceedings of the National Academy of Science of the U.S.*, v. 89, December 1992, p. 11658.

Bibliography

Alcamo, I., *DNA Technology The Awesome Skill*, Dubuque IA: Wm. C. Brown Publishers, 1996.

Bigbee, D., Tanton, R., and Ferrara, P., Implementation of DNA Analysis in American Crime Laboratories, *The Police Chief*, October, 1989.

Brookfield, J., Chakraborty, R., and Weir, B., DNA profiling on trial, *Nature*, v. 369, June 2, 1994.

Brown, J., DNA Analysis: A Significant Tool for Law Enforcement, *The Police Chief*, March 1994.

Burke, T., and Rowe, W., DNA Analysis: The Challenge for Police, *The Police Chief*, October 1989.

Chakraborty, R., and Kidd, K., The Utility of DNA Typing in Forensic Work, *Science*, v. 254, December 20, 1991.

Cohen, J., Lynch, M., Taylor, C., Green, P., Lander, E., Devlin, B., Risch, N., and Roeder, K., Forensic DNA Tests and Hardy-Weinberg Equilibrium, *Science*, v. 253, August 30, 1991.

Devlin, B., Risch, N., and Roeder, K., No Excess of Homozygosity at Loci Used for DNA Fingerprinting, *Science*, v. 249, September 21, 1990.

Good, I., When Batterer Becomes Murderer, *Nature*, v. 381, June 6, 1996.

Hagelberg, E., Gray, I. and Jeffreys, A., Identification of the skeletal remains of a murder victim by DNA analysis, *Nature*, v. 352, August 1, 1991.

Harding, R., Boyce, A., and Clegg, J., The Evolution of Tandemly Repetitive DNA: Recombination Rules, *Genetics*, v. 132, November 1992.

Howlett, R., DNA forensics and the FBI, *Nature*, v. 341, September 21, 1989.

Koshland Jr., D., DNA Fingerpinting and Eyewitness Testimony, *Science*, v. 256, May 1, 1992.

Koshland Jr., D., The DNA Fingerprint Story (Continued), *Science*, v. 265, August 19, 1994.

Landers, R., Solving Crimes With Genetic Fingerprinting, *Editorial Research Reports*, v. 1, n24, June 30, 1989.

Lehrman, S., UK to set up DNA database of criminals, while US may regulate DNA testing laboratories, *Nature*, v. 370, August 25, 1994.

Lewontin, R., and Hartl, D., Forensic DNA typing dispute, *Nature*, v. 372, December 1, 1994.

Lewontin, R., and Hartl, D., Population Genetics in Forensic DNA Typing, *Science*, v. 254, December 20, 1991.

Macilwain, C., Ceiling principle "not needed" in DNA cases, *Nature*, v. 381, May 9, 1996.

Mervis, J., Academy's About-Face on Forensic DNA, *Science*, v. 272, May 10, 1996.

Nowak, R., Forensic DNA Goes to Court With O.J., *Science*, v. 265, September 2, 1994.

Risch, N., and Devlin, B., On the Probability of Matching DNA Fingerprints, *Science*, v. 255, February 7, 1992.

Roberts, L., Fight Erupts Over DNA Fingerprinting, *Science*, v. 254, December 20, 1991.

Roberts, L., Science in Court: A Culture Clash, *Science*, v. 257, August 7, 1992.

Shapiro, M., Imprints on DNA fingerprints, *Nature*, v. 353, September 12, 1991.

Weir, B., Independence of VNTR Alleles Defined as Fixed Bins, *Genetics*, v. 130, April 1992.

Weir, B., Population Genetics in the Forensic DNA Debate, *Proceedings of the National Academy of Science of the U.S.A.*, v. 89, December 1992.